Review Report: Plant Ontology

Reviewers: Bjoern Peters, Melissa Haendel, Janna Hastings

Version of ontology reviewed:
http://palea.cgrb.oregonstate.edu/viewsvn/Poc/tags/live/plant_ontology.obo
Release #19 December, 2012 was used (See http://wiki.plantontology.org/index.php/Release_19).

Domain of ontology:
Plant anatomy, morphology and growth and development and their association to plant genomics data for all plants.

Ontology contact: Laurel Cooper, cooperl@science.oregonstate.edu

Our specific recommendations are contained in the report below. Thank you for participating in the OBO Foundry review process.

Summary of Review Findings

In general, we found that the Plant Ontology is a good ontology, largely following the principles as laid down by the Foundry. We have some specific recommendations for improvements but found no major significant shortfalls in the offering. The following table summarises the findings of this review on a per principle basis, and detailed discussion of each point follows.

Principle	Level
1: Open	Some improvement desirable
2: Common format	Pass
3: Foundry URIs	Pass
4: Versioning	Pass
5: Clearly delineated content	Some improvement desirable
6: Textual definitions	Some improvement desirable
7: Relations	Some improvement desirable
8: Documentation	Pass
9: Plurality of users	Pass
10: Collaboration	Pass
11: Single locus of authority	Pass
12: Naming conventions	Pass
16: Maintenance	Pass

Principle 1: The Ontology is Open

✓**Some improvement desirable**

The self-review states that the PO is released under a Creative Commons license. However, it is not specified *which* Creative Commons license has been chosen. The ontology file does not contain a statement about the license or the policy for reuse of the ontology. However, the information is available on the website e.g. at http://plantontology.org/docs/otherdocs/principles_rationales.html: *Plant Ontology by Plantontology.org is licensed under a Creative Commons Attribution-NoDerivs 3.0 United States License.*

We recommend that the exact license chosen be included in an additional ontology annotation in the ontology file. According to the OBO Foundry principle 1 documentation, this should make use of two annotations using the following properties:

dc:license
rdfs:comment

The license type goes in the dc:license field. The mode of attribution goes in the rdfs:comment field. While Creative Commons 3.0 CC-by license is recommended, this is not mandatory and therefore the chosen attribution-noderivs license is also compatible with the Open principle. However, we urge you to consider weakening the license to one that encourages sharing and reuse without constraint to further the adoption by the community and the value of the resource for the community.

Principle 2: Common Format

✓**Pass**

The ontology is available in both OBO and OWL formats. It should be noted that the OBO form of the ontology contains semantics which are not yet available in the OWL form (i.e. the treat-xrefs-as-isa and treat-xrefs-as-genus-differentiae constructs). These are supported by the Oort release manager and can be instantiated in OWL versions of the ontology. See https://code.google.com/p/owltools/wiki/OortOptions

Principle 3: Foundry-style URIs

✓**Pass**

The ontology uses numeric Foundry URIs.

Principle 4: Versioning

✓**Pass**

The ontology uses date stamps for versioning. These are reflected in the OBO and OWL versions appropriately. We note that the PO version in OntoBee was not the same as their latest release described on their website and available from the OBO Foundry download page at the time of doing the review, however, this seemed to have been a problem with the OntoBee updating mechanism rather than with the PO itself.

Principle 5: Clearly Delineated Content

✓Some improvement desirable

There are several sub-issues that are contained in this principle. Firstly, the ontology needs to have a clearly delineated scope as expressed in a statement of scope. PO does have a clearly delineated scope, and this scope does not substantially duplicate the scope of any other OBO Library ontology. Secondly, the ontology needs to have a clear upper level with root terms having definitions that indicate the content of the sub-ontologies included in the ontology. Here, we have some concerns as detailed below. Thirdly, the ontology should not contain terminology that overlaps with other OBO ontologies except where content has been imported for reuse in specific contexts appropriate to the domain of the ontology. Finally, the ontology should provide adequate coverage with respect to its stated domain so as to justify its use in applications.

Concerns about the upper levels of the plant ontology:

- The definition of the root term 'plant anatomical entity' appears intended to be a plant specific subclass of the CARO term 'anatomical entity' and is defined as "An anatomical entity that is or was part of a plant." The CARO class states that an anatomical entity is part of or located in an organism (or virus or viroid in the upcoming release of CARO), however here you include things that were once part of a plant. In OBI or ERO, such entities would be considered specimens or samples. Can you explain why this is needed? If it is justified, there is some cross-ontology coordination that should be performed.
- There is a need to define the high-level distinction between plant structure vs. portion of plant substance better: currently 'plant structure' is defined as anatomical entity that is or was part of a plant; but substances are also (or were also) part of a plant. This is likely because these are intended to be subclasses of the CARO classes, that have more specific differentia, however, there is no mireot or subclass assertion to these classes other than in the comment field. Explicit use of the CARO classes will help differentiate these top-level nodes, and inclusion of their logical class definitions can help with error-checking.
- If you consider the above point on the definition of 'anatomical entity' referring to in vivo entities, you could consider moving 'in vitro plant structure' to be its own root outside anatomical entity, as it would technically be a derived structure and not part of or located within an organism (or are most of these simple plants cultured in vitro?). Efforts have been made recently in the CL and OBI to represent such things (e.g. "in environment" vs. "ex environment", you could potentially leverage some of these design patterns and/or coordinate with them.

Other concerns from somewhat random sampling:

- 'cultured plant embryo' has two asserted superclasses. While asserted multiple inheritance is debatable in terms of ontology maintenance or ontological rigor (GO has a lot of this), in this case the example ensues because of a lack of logical class expression on this class. For example, this class could instead leverage a class restriction that states 'plant embryo' and is_specified_output of some in vitro plant culturing. In this way, you would infer the second parent, 'in vitro plant structure' (dependent on how you define things as per above). This

would mean higher quality axioms, easier ontology maintenance, and interoperability with other ontologies such as OBI.

Principle 6: Textual Definitions

✓Some improvement desirable

We applaud the fact that the ontology does contain text definitions for all terms included in the ontology (see Appendix: Metrics) and for its extensive synonyms in numerous languages. However, we note that some of the text definitions do not exactly match the formal definitions as given in equivalence class or subclass axioms (specific examples contained below) and furthermore that not all of the textual definitions are as clear as they could be e.g. by following the genus differentiae pattern (examples below). We advise continued evolution of the textual definitions contained in the ontology towards greater clarity for the benefit of all users, and gradual inclusion of a greater number of equivalent class definitions.

Specific points:

- Definitions such as that for 'microgametophyte', referencing IDs within the text, should be converted into logical definitions.
- 'pollen' is textually defined as located_in some 'pollen sac' and logically it has the relationship part_of some pollen sac. This is an inconsistency with respect to the normal usage of these two relationships.

Principle 7: Use of Relations

✓Some improvement desirable

For the most part, the PO uses RO relations. The relation that is PO specific is provided with adequate documentation in the form of a definition and comment. Some relations are, however, defined in the PO that are within RO or BFO scope and yet do not have appropriate cross-references/Mireoted URIs. These include 'has_participant', 'located_in' and 'participates_in'. This should be addressed, and full inclusion of such properties will facilitate the use of their logical properties for error checking.

Specific points:

- derives_by_manipulation_from is described as a subproperty of derives_from, whereby there has been a human processing step. This appears to be some kind of shortcut but isn't yet logically defined as such. We would urge PO to work with OBI to leverage the work they have done in defining organismally derived entities, whereby such entities are the outcome of a 'material processing' process, has_specified_output the processed entity, and are derived_from the parent anatomical entity (for example see 'urine specimen').
- Plant developmental stage classes do not yet have precedes relations for temporal ordering.
- Consider using the stage relations (for example, existence_starts_during) from Uberon and similar to the zebrafish anatomy ontology, to relate organism developmental stages to the organismal entities at those stages. For example, 'plant embryo' could be defined as 'whole

plant' and (existence_ends_with some 'embryo stage') and (existence_starts_with some 'embryo stage') rather than use of the participates relation.

Principle 8: Documentation

✓Pass

Extrinsic: The ontology is well documented with an extensive website and several publications.

Intrinsic: The ontology has high quality text definitions for every term.

Specific points:

- There is some conflation in the use of annotation properties. A lot of information is placed in the comments field, including examples of usage, subclass assertions, and stage or taxon specific considerations. You may wish to break these out into more specific annotation properties. See IAO-core metadata and/or Uberon for some options, and please feedback to IAO if you desire a standard annotation property that isn't yet available.
- In some cases, definitions are attributed to "POC:curators". We'd urge you to consider use of specific persons or sources for attribution, and as the Foundry moves towards the use of ORCID IDs, to include these as part of your attribution strategy.
- While we applaud the reference of tracker items directly in the ontology, these items can be included as URLs such that the link is easily made directly from the ontology, for example see 'leptome' to "OBO_SF_PO:3295055". Similarly with the PO_REFs- please make these URLs or imports so that they can be easily located. It is likely that a new standardized annotation property for this purpose should be created and used across the Foundry.

Principle 9: Plurality of Users

✓Pass

The ontology contains a plurality of users as evidenced by the wealth of citations to the main publications in which the project has been presented (three top papers each cited >50 times).

Principle 10: Collaboration

✓Pass

The ontology has established many important collaborations and serves as a hub for the exchange of data and annotations in the domain as evidenced by the discussion contained in the paper Plant Ontology paper (Comparative and Functional Genomics 6 (7-8) 2005, p 388-397) and by the many citations. The PO developers have furthermore also participated in conferences and attended meetings hosted by developers of ontologies in other, related domains. Finally, the PO has offered a tracker to encourage community participation. The SourceForge tracker has 537 requests in total and there is evidence of continuous and recent responsiveness by the PO team to these requests.

Principle 11: Single Locus of Authority

✓**Pass**

The PO has specified a single locus of authority.

Principle 12: Naming Conventions

✓**Pass**

The PO uses singular nouns for naming the entities contained in the ontology. Names are unambiguous. Some terms, such as 'fruit', have an exceptionally high number of alternative IDs. Use of alternative IDs should be minimised wherever possible. However, it is understood that this scenario can arise for valid reasons having to do with legacy, therefore it is not specifically penalised.

Principle 16: Maintenance

✓**Pass**

By their own admission, the PO is committed to the maintenance of the ontology in the light of scientific advance. The tracker and SVN repository show a history of ongoing maintenance that serves as sufficient grounds for optimism about continued performance into the future.

Appendix: Ontology Statistics

Statistics from BioPortal:

Number of classes:	1644
Number of individuals:	0
Number of properties:	10
Maximum depth:	19
Maximum number of siblings:	61
Average number of siblings:	1
Classes with a single subclass:	190
Classes with more than 25 subclasses:	5
Classes with no definition:	0

Ontology Metrics from Protégé:

Axiom count	20597
Logical axiom count	2616
Class count	1644
Object property count	8
Data property count	0
Individual count	0
DL expressivity	SR
SubClassOf axiom count	2498
EquivalentClasses axiom count	78
DisjointClasses axiom count	36
GCI count	0
Hidden GCI count	78
AnnotationAssertion axiom count	15924